

Using Mobile Data for Understanding Population Movement and Disease Transmission during Covid-19 Outbreak in the Nordics

Osama Mansour
Lund University
osama.mansour@ics.lu.se

Miranda Kajtazi
Lund University
miranda.kajtazi@ics.lu.se

Ahmad Ghazawneh
Halmstad University
ahmad.ghazawneh@hh.se

Abstract

This study investigates the use of mobile data to understand patterns of population movements and disease transmission during the Covid-19 outbreak. It also focuses on understanding the implications of using this data for individual privacy. Using a mixed methods approach, we present 10 rich qualitative interviews and 412 survey responses from participants across the Nordics. Our novel results show that the use of mobile data can be characterized by two main categories: validation data and complementary data. We also identify five implications for practice: sharing resources and expertise between health agencies and telecom companies; extended collaboration with multiple network operators; cross-disciplinary collaboration among multiple parties; developing data and privacy guidelines; and developing novel methods and tools to address the trade-off between maintaining individual privacy and obtaining detailed information from mobile data. These implications may inform immediate and future actions to prepare for, mitigate, and control the spread of infectious diseases using mobile data. They also show privacy-driven limitations of mobile data in terms of data accuracy, richness, and scope.

1. Introduction

Modern digital technologies such as smartphones, the Internet of Things (IoT), and Artificial Intelligence (AI) are ubiquitous in our daily lives and generate massive amounts of real-time data, so-called big data [1,2,3]. However, their application to understand disease transmission during the first pandemic in the era of big data and the spread of the Coronavirus disease (Covid-19), has only recently been acknowledged [4]. These technologies are major sources of data and provide detailed information about spatial and temporal dynamics [1,5] that can be critical to understand population movements in disease outbreaks.

Nowadays, in the midst of a global effort to combat the spread of Covid-19, governments and public health authorities in various countries are using mobile data to analyze population movements in order to monitor, prepare for, and control the transmission of the disease. One of the most important parameters in

monitoring and eventually controlling pandemics is transmissibility, that is person-to-person transmission [6]. This emphasis on human movement was already acknowledged since the time of the Black Death when authorities were surveilling the movement of people and enforced quarantine time for anyone suspected to show symptoms of an infectious disease. [7].

Public health surveillance is said to provide early and timely data about key epidemiological, clinical, and virological characteristics of pandemic viruses [8,9,10]. Thacker and Stroup [11] pointed out that modern era disease surveillance was initiated in the 1950s. Usually, public health surveillance involves collecting data from multiple sources, such as, hospitals, clinics, public health centres, public health labs, and individual physicians, which are then stored and shared using a national public health information system [11,12]. This kind of surveillance has been described as fragmented mainly because of poor coordination among surveillance systems, which often results in ineffective data [8,11]. For instance, Briand et al. [11] discussed global surveillance information systems appropriated by the World Health Organization [13] such as the Global Influenza Surveillance Network (GISN). This system is designed to report geographically-based influenza-like activity using a web-based electronic data reporting, query, and mapping system. They explained that such systems entail challenges including timeliness of information, lack of surveillance capacities to monitor an outbreak over an extended period of time, and accuracy problems. At times of disease outbreaks, such challenges may prove critical when there is an urgent need for timely and accurate data to monitor the rapid transmission of highly infectious diseases [8,11,14].

Current data tools enable the collection and analysis of massive amounts of data from multiple sources and offer novel opportunities for real-time detection and monitoring of disease outbreaks like Covid-19. Mobile phones are major sources for big data especially that it is possible for every action taken by a mobile user to be traced and geo-tagged [1]. For instance, many mobile communication companies offer location-based services to provide data for businesses about where customers buy things. These services use real-time data about customers which allow businesses to understand not

only customers' locations, but also how their needs and interests change [1], making it possible to do spatio-temporal behaviour analysis of an individual's every move [15,16]. In the context of disease outbreaks, there is already evidence in literature that mobile data is useful in understanding population movement patterns by making predictions of vulnerable geographic areas and developments in infection rates [5,17,18,19]. Also, the potential of using real-time data from mobile communications about spatio-temporal dynamics of individuals and populations may help with rapid detection and response which facilitates preparedness measures by public health authorities. As Covid-19 continues to spread around the world, countries like China, South Korea, Singapore, and Israel [34] have tasked their intelligence units to use surveillance tracing [33] for tracking new Covid-19 infections. For instance, Israel's Shin Bet used its cellphone localization intelligence capabilities to identify individuals who have been in contact with others who tested positive for Covid-19. Other countries like Sweden chose a different approach to use mobile data where no privacy-invasive surveillance was applied. Telecom companies in collaboration with health authorities to understand general movement patterns using some sort of proximity surveillance [33] and data aggregation [16].

While the potential of mobile data may be promising, it also comes with limitations regarding data accuracy and privacy [5,18]. It is also not yet clear how mobile data may be useful in providing details of population movement patterns (i.e., how, when, and where people move) [5,17] that can be used for understanding disease transmission dynamics [5]. As such, there is a need for a nuanced understanding of spatial and temporal factors of population movement which is evident to be a vehicle for disease transmission [5,9,19,20]. The near absence of peer-reviewed research on using mobile data in disease outbreaks like Covid-19 prompts an urgent need for knowledge contributions [22]. The aim of the current study is therefore twofold: first, to develop an understanding of using mobile data to analyse population movement patterns to monitor the spread of Covid-19; second, to identify implications of using this data for understanding disease transmission dynamics as well individual privacy. These aims are achieved by seeking answers to the following questions: *How can mobile data be used to understand the transmission of Covid-19? What are the implications of using mobile data for individual privacy?*

2. Human movement, mobile data, and privacy in pandemics

It is well established in epidemiological literature that human movement and international travel to

endemic places often resulted in imported cases and disease outbreaks [9,17,19,20]. Jiang & Luo [17], for instance, investigated the influence of human movement on the transmission of Covid-19 and concluded that it is a key driver for the rapid transmission of the disease. In absence of effective prevention and treatment during disease transmission of an emerging infectious disease such as Covid-19, public health preparedness is the most important measure to minimize mortality rates. Developing models and tools to help in predicting the geographic spread and timing of epidemics is critical for spatio-temporal targeted interventions and epidemic preparedness [19].

Epidemiologists often use mathematical and computational models to study spatio-temporal dynamics of disease outbreaks [23]. The purpose of using these models is to forecast outbreak dynamics such as peak time, peak height, and magnitude in order to inform policy makers and public health practitioners regarding preparedness measures and the allocation of resources [23]. However, making reliable forecasts using these models proved to be challenging due to the difficulty of making timely forecasts [8]. In this respect, contact tracing has been used as a standard strategy for tackling new or re-emerging infections especially when the number of infectious cases is low, whether in the beginning of a pandemic or in its later stages [24]. While contact tracing is deemed an effective measure to break transmission chains [24, 25], the process of contacting infected individuals to name their contactees for locating unidentified cases can be slow resulting in more disease transmission. This is because the tools used for contact tracing, also called partner notification, such as face-to-face interviews do not produce real-time data about spatio-temporal dynamics of human movement. Eames & Keeling [24] explained that accurate modelling of contact tracing requires explicit information about disease-transmission pathways from each individual, and hence the network of contacts.

More recently, many countries around the world turned into using digital contact tracing based on mobile apps [32, 33] in countering Covid-19. Urbaczewski & Lee [34] mentioned that as many as 80 contact tracing apps were used in 50 countries just a few months after the first cases of Covid-19 were detected. Mobile-based contact tracing uses cell phone network data, global positioning system (GPS), and Wi-Fi signals to identify geolocations and measure the proximity to infected individuals. Riemer et al. [33] identified three different contact tracing approaches including traditional manual tracing, surveillance tracing, often by governments, and uses surveillance technologies such as CCTV cameras and facial recognition, and proximity tracing which is based on collecting proximity data when contact, rather than tracing, between people take place. Countries that

used surveillance tracing enjoyed varying degrees of success such as Israel, Singapore, and China [34] in terms of reducing Covid-19 cases but faced widespread criticism for privacy and even human rights violations. Often these governments use laws (e.g., Shin Bet Data Collection law in Israel) which allows them to collect identifiable data about individuals. In Singapore the prime minister asked citizens to install a tracing app called TraceTogether on their personal mobile phones.

The use of mobile data in disease outbreaks might contribute into traditional public health surveillance by offering explicit and critical information. Cinnamon et al. [18] discussed that mobile communications produce massive longitudinal datasets or ‘mobile big data’, especially Call Detail Records (CDRs) which include data about communication time, a unique caller ID, and the location of the cellular tower or base station through which the communication is routed. They argued that CDR data has the potential to illuminate spatio-temporal dynamics of populations in real-time, which can be useful for preparedness purposes at times of infectious disease outbreaks. This can be seen in [19], a study which develops an epidemiological model of disease transmission during the dengue epidemic in Pakistan using mobile data. They explained that mobile data offer direct measures of human aggregation and movement patterns that allow for accurate mapping of changing patterns of vulnerability in a country.

However, while mobile communications may produce massive amounts of real-time data, their application and use may be limited in practice. Data accuracy is a major problem because of privacy and security measures applied by telecom operators such as data aggregation and anonymization [18]. There are also other limitations caused by the lack of guidelines to use mobile data that can be acceptable by customers and useful to health agencies [18,19]. In general, preserving the accuracy of data while maintaining individuals’ privacy remains one of the key challenges of the big data era [26]. Acquisti et al. [27] suggested that regulatory interventions are needed to address the challenges of a rapidly evolving privacy landscape. The vulnerability with privacy remains alarming because one technical negligible act can lead to revealing details about users’ identity, which can infringe privacy rights, such as those created by the GDPR. The analysis of users’ locations, for instance, has shown that people tend to develop a negative attitude towards it and adjust their privacy settings to gain a sense of control over their data [28]. The potential loss of control over personal data during the Covid-19 outbreak has led Human Rights Watch to be concerned about countries that use mobile data to collect location data due to risks of revealing sensitive individual information.

3. Research method and settings

3.1. Use of mobile data during Covid-19

The empirical context for the current study is Sweden and Norway where public health agencies initiated collaborations with telecom companies to use mobile data for understanding Covid-19 transmission:

The collaboration between the Swedish Public Health Agency (FHM) and telecom operator Telia: FHM is the national public health agency in Sweden which is responsible for “disseminating scientifically based knowledge to promote health and prevent disease and injury”. One main task of FHM is to protect against communicable diseases on a national level. On April 8, 2020, FHM announced on its official website a collaboration with the largest telecom company in the Nordics, Telia, to use mobile data in understanding population movement to counter the Spread of Covid-19. The statement said that FHM will use mobile data to study how population movement affect the virus spread as well as provide a basis for enacting measures and assessing existing measures. To obtain this data, FHM has access to a dashboard which is developed based on an existing service owned by Telia, called Crowd Insights. The raw data is the signals between mobile phones and base stations which can give an approximate location of the mobile phone at a certain time. Patterns of how groups of more than 5 people move during 24-hour time chunks are collected, extrapolated to represent the whole population, and visualized in Crowd Insights.

The collaboration between the Norwegian Institute of Public Health (NIPH) and Telenor: NIPH is a national competence institution placed directly under the ministry of health and care services which has the responsibility for controlling infectious diseases. In January 2020, and as the Covid-19 disease was spreading globally, a senior data scientist who leads a small research center at Telenor approached his colleagues at Telenor and NIPH and suggested an initiative to explore the potential of using mobile data to understand the transmission of Covid-19 in Norway. The aim was to develop a regional model focused on municipalities to predict the transmission of Covid-19. Since then, Telenor provided NIPH with daily updates and aggregated data sets that are processed using a data analytics system called Spark which shows a quantification of patterns of travels and movements across different regions in the country. Similar to Telia, the data from Telenor is aggregated and anonymized but aggregation is done based on a group of 20 people. The data is eventually shared by Telenor into a secure server on a daily basis which NIPH can access and download.

3.2. Participant selection and data collection

The empirical data were collected using qualitative interviews and a quantitative survey. The interviews aimed at collecting data from experts involved in using mobile data during the Covid-19 outbreak, while the survey was focused on collecting data from people affected by the outbreak to get individual perspectives on privacy issues. The aim was to triangulate the data to get various perspectives from two different groups of participants and to draw out better empirical insights.

The qualitative empirical data were collected from participants employed by FHM, Telia Sweden, NIPH, and Telenor Norway. The selection of these participants was done using a combination of purposive and snowball sampling. The sample included participants specialized in three main areas: data analytics, data privacy, and epidemiological models. All of them are experts, and most hold a PhD, in their respective areas, and have been working for several years in their organizations. Our first communication was made with FHM in Sweden through sending an email to inquire about their collaboration with Telia. This resulted in access to our ‘first contact’ who is a data analyst and was responsible for the collaboration with Telia. To expand the scope of our study, we took the same approach with Telenor Norway and got in touch with a senior data scientist who initiated the collaboration with NIPH. Our first contacts were asked to recommend others who can join our study and the sample has then gradually increased up to a total of 10 participants as shown in Table 1. below.

#	Position	Organization	Country	Gender
1	Data analyst	FHM	Sweden	M
2	Head of data insights	Telia	Sweden	M
3	Senior data scientist	Telenor	Norway	M
4	Head of Infectious Disease Modelling	NIPH	Norway	F
5	Director of privacy	Telenor	Norway	F
6	Data protection officer	Telia	Sweden/ Finland	F
7	Head of analysis	FHM	Sweden	F
8	Professor 1, 2, 3	Lund Univ.	Sweden	F/M/F

The primary vehicle for the qualitative data collection was the interview method [30]. An interview protocol was developed to guide the conversation with the participants, which was based on a review of literature on using mobile data in pandemics [5, 17, 18] privacy issues [16, 27]. This protocol was structured into four main parts including general info about the interviewee and the collaboration between the agency and the telecom operator, technology for mobile data collection and analysis, data-driven health measures, and privacy and legal issues. While each part contained predefined questions, the interviews were largely semi-

structured, and emphasis was placed on following up on curious answers rather than strictly sticking to protocol.

All participants were interviewed via an online medium. The geographical distribution of the participants as well as the restrictions on face-to-face meetings during the Covid-19 outbreak were major reasons for choosing an online medium, Zoom and Microsoft Teams, to conduct the interviews. The interviews were conducted between May and June 2020, and the range of interviewing time was between 50 and 80 minutes. Ethical issues were communicated and permission to record the interviews was obtained from all the participants, except one (read more below). All interviews were eventually transcribed verbatim, and the participants were sent copies of our findings for validity checks [31]. It should be noted that a focus group interview involving all authors of the paper was conducted with two professors from Lund University, Sweden who lead a project on using a mobile contact tracing app for studying Covid-19 transmission. An additional interview with a third professor involved in the project took place at a later time and tackled the third part of the interview guide which is focused on privacy matters. Permission to record was not granted by this participant, but a confirmed list of notes from the interview was approved. The focus group and the additional interview with the third professor were aimed at getting a general understanding of how a mobile app can be used to collect data about Covid-19.

3.3. Survey development and demographics

The survey was aimed at collecting a user perspective on using mobile data with emphasis on aspects such as trust, population movement, privacy issues, and public health. The questions developed for the survey were based on both relevant literature as well as insights from the interviews as our primary source for analyses. The interview data inspired our attention for using sensitive terms in the survey that could be misinterpreted by our respondents such as avoiding to use terms like “monitoring” and “surveillance” and choosing to use “population” rather than “individual”. The survey was designed by the authors, primarily focusing on our interview guide, where a back-and-forth activity between the authors led to formulate the survey in Qualtrics, further distributed by Prolific. Each participant was reimbursed with a recommended amount by Prolific to respond to our survey. The distribution of the survey focused on targeting a specific population of our interest that led to custom pre-screening, with nationality being of key importance, since we aimed at collecting data from citizens in the Nordics only. Out of 850 matching participants, we had a response rate of 48.47% with 412 recorded responses.

We kept the survey active until we reached 400+ responses, then when we realized that the rest of the matching profiles were not interested to join our survey, before we closed the survey and agreed that the number of respondents in relation to the number of questions was optimal for our intended quantitative analyses. Each question had to be answered before moving to the next question/page. Due to limited space, the full survey can be accessed via an active link in the footnote below.¹

Table 2 below shows the demographics for highlighting our controls, however, controls like gender in particular were not key to witness any differences, but it shows that the majority of our Prolific matching participants were male and yet we did not consider this to be an issue and that it would make an impact in our analysis concerning the intention of our study. The reason why we chose nationality as a factor, whilst not considering country of residence, is because we realized that certain respondents (e.g., an exchange student) may have resided in a country for a short time (e.g., months) which makes it harder to generally reflect on the work done by the public health agency for instance.

Demographics			
Control variable	Variable	Frequency	%
Gender	Female	116	28.22%
	Male	288	70.07%
	Nonbinary	7	1.70%
Age	Under 20	66	16.06%
	21-30	204	49.64%
	31-40	101	24.57%
	41-50	21	5.11%
	51-60	15	3.65%
	61-70	4	0.97%
	70+	0	0
Education	High school	193	46.96%
	Undergraduate	114	27.74%
	Graduate	104	25.30%
Nationality	Swedish	147	35.77%
	Danish	83	20.19%
	Finnish	114	27.74%
	Norwegian	50	12.17%
	Icelandic	15	3.65%

3.4. Data analysis

Our data analysis process can be described as analytic induction [31]. Patton [31] describes analytic induction as an analysis technique where one starts analyzing the data deductively followed by an inductive analysis of the data to generate insights. So, we first used the literature deductively to develop our interview and survey questions as well as conduct initial reading and understanding of the interview data. After transcribing the data, all authors read the interview transcripts separately. An author would first read the transcript thoroughly to develop an overall understanding of the interview. This step was driven by both previous understandings from literature (deduction) as well as an emergent understanding originating from the data itself (induction). Then,

meaningful data segments (i.e. quotes) from each transcript were identified. The identification of data segments was essentially an interpretive activity which involved making sense of the text as well as drawing meaning out of it which is either literature-based or emergent. Each author applied this interpretive, sense-making activity in all interview transcripts. Eventually, a number of relevant data segments were identified in each transcript and coded using descriptive terms (e.g., privacy) for further cross-transcript, all-author analysis.

The next step in the analysis process was done in a meeting involving all three authors of the paper where each transcript was discussed. The authors discussed all data segments identified earlier by them in each transcript and then filtered out the most relevant and meaningful segments to develop higher-level themes and patterns. This discussion was critical to develop a shared and consistent understanding of the data by all authors as well as selecting key themes and patterns together with strong empirical data segments that support them. These selected data segments form the basis of our empirical findings which are presented below. For survey data, the analysis was performed using descriptive statistics to develop a general overview of issues related to analyzing population movement in terms of benefits, trust and privacy. We then used the key control variables (gender, age and nationality) to see if they impacted the views differently in correlation to benefits, trust and privacy matters.

4. Findings

4.1. Mobile data aggregation and the development of population movement patterns

Mobile data processed by mobile network operators are usually generated from raw cellular signals between customers' mobile phones and base stations or cell towers. This data includes a variety of details including the location, since each individual phone connects to a base station whenever a phone user makes a call or sends a text. Due to data sensitivity, strict privacy regulations are enforced by telecom companies to ensure no individual is identified. Data aggregation, performed using data analytics tools and applications, is a key measure for ensuring user privacy and anonymity at both telecom companies discussed in this paper. Mobile data referred to here is therefore aggregated data. The Director of Privacy at Telenor Norway explained data aggregation: "...aggregation is basically a part of anonymization measures, so it's one measure to receive like a good outcome of anonymization and to

¹ https://docs.google.com/document/d/1IMk947-omBxn_PwNiMk6Cfqdtbrhu_zFW075oJIGSo/edit?usp=sharing

basically end up with a good a anonymization. So, the lower the level or the number of people or the group of people, the easier it will be to identify one person out of these...so it doesn't not have to be possible to re-identify based on all the datasets available."

The aim from data aggregation in the context of Covid-19 outbreak is to develop generic high-level patterns of population movements. These patterns are operationalized into so-called "origin-destination" matrices that can help to show how many people went from one point to another (i.e. from one municipality to another), or how many people are present in a specific location. The Senior Data Scientist at Telenor explained data aggregation and the development of matrices of movement patterns as follows: *"...the individual, personal location data of all customers are processed and aggregated into origin-destination matrices. When we talk about travel patterns, one should have in mind origin-destination matrices, because this is the operationalization of travel pattern."* At Telia Sweden, similar origin-destination matrices are developed where data aggregation is also applied to ensure anonymity. The Head of Data Insights explained how data aggregation is done and the level of granularity that is possible to generate from the data, he said: *"...our data is anonymous and aggregated so it's not very granular, we can go down to 500*500 meters at best, and this is only in highly populated areas, like if you go out in the countryside, due to anonymization and aggregation rules you might look at , you know, 10*18 meter grids. So, we cannot see if people are very close, but we can see how many are in this 500*500-meter square."* Eventually, this aggregated data is visualized using tools like Tableau into an interactive dashboard that contains rich and dynamic insights as well as analytical models. The content in the dashboard is the main outcome of data aggregation for FHM in Sweden. They have access to this dashboard from Telia Sweden where they can see maps with travel patterns across the country.

4.2. Potential use and limitations of mobile data in the Covid-19 outbreak

The Senior Data Scientist from Telenor Norway described the potential of mobile data using a metaphor of steering a ship in the ocean. He explained that the ability to make predictions two or three weeks in advance is important to know whether the ship is heading towards a wreckage or smooth water. In the context of the Covid-19 outbreak, he stated that the aim is to help NIPH *"build a regional model based on municipalities in Norway predicting coronavirus spread"* so they can predict when and where the spread might be happening. In Sweden, a data expert from FHM explained that one of the main aims why the

agency wanted to look into mobile data is to see if people actually commit to the recommendations. He said: *"we try to encourage people not to travel during the holidays and stay at home. You know just to get this confirmation that people have actually listened and heard what we say..."*. The Head of Data Insights at Telia Sweden shared an elaborate view of the potential of using mobile data, he said: *"It is really two parts. It is about understanding the effect of the recommendations to see if citizens are adhering to recommendations...And in the second part is using the dataset, this data source as an input in epidemiological models."*

In this respect, while survey data show that more than half of our respondents believe that understanding population movements using mobile data is beneficial for health agencies, many of them do not see how this can be beneficial for them at the individual or community levels. We also noticed that the majority of our survey respondents, more than 60% (248), reported that understanding population movements is beneficial to stop the spread of Covid-19, while around 17% (74) did not consider that it is beneficial. The respondents who did not consider any benefits were evenly represented across the Nordics, and almost all of them seemed to be worried about their anonymity. Further, using mobile data as a potential data source for epidemiological modelling is a novel area of interest for epidemiologists since it is not something they used before in a disease outbreak. The Head of Methods Development at NIPH provided an epidemiological perspective on the potential role and importance of mobile data, she said: *"If I should say to you when these mobile data are mainly important, it is in the phase where there is not so large level of transmission occurring because at a certain point we'll have a lot of local transmission and therefore, transport between places becomes less important."* Similarly, the Head of Analysis at FHM, further elaborated: *"It is not a question of being better prepared if we have access to more mobile data. It is knowing the quality of the data, what can it be useful for, how can we transfer these data into some kind of knowledge or predictions."* Despite the lack of clarity on how mobile data can be incorporated in epidemiological work, the Head of Data Insights at Telia Sweden believe that the kind of insights they generate using their algorithms from mobile data can be useful to epidemiological work in two areas. He explained: *"...the first one is, in these epidemiological models you need to estimate the amount of contact people have. It makes sense, right? Viruses cannot transmit disease unless you're in contact with someone, so they typically have some sort of value that approximates contact...the second part is ... in epidemiological models you want to model the special*

aspect...so these are the two parts: it is the estimation of probability contact and then migration patterns.”

The Head of Analysis at FHM, discussed how the data is currently used during the Covid-19 outbreak and the limitations of using such data, she stressed: *“If we could have data, which age groups and from which regions the travelers come from, it will be useful for understanding and validating the (disease) dynamics ...As it is right now, since it is aggregated, we have it as a complement, as one piece of knowledge to our knowledge”*. There are also other issues that might limit the potential of mobile data in disease outbreaks due to privacy concerns. For instance, the Director of Privacy at Telenor Norway explained that when aggregated datasets do not conform to the rules, the algorithm used to analyze the data and develop movement patterns will delete the information, which is in fact a censoring process. She explained: *“... we do aggregate it (movement patterns) to 20 people so that we can really make sure that the data is actually anonymized, because like you might have an area somewhere where very little people in Norway or where a specific point of time there were only one person and then we basically without mentioning which this person this is but say, well, at this point of time there was some person, this would not be anonymous data so all datasets where are less than 20 people will be deleted.”* Another limitation observed in the data was related to collecting “real-time” data. The Head of Data Insights at Telia Sweden discussed legal and ethical concerns for why they won’t and can’t collect real-time data. He said: *“We do not do real-time, like same-day information, because of the legal anonymization process, and also the ethical consideration, because knowing where large groups of people are, there could be potential negative uses from that. So it is both legal but also ethical safeguard that is 48 hours delayed.”* Further, the Data Protection Officer at Telia Sweden identified privacy-related factors that affect how mobile data are used. She said: *“...we need to have quite a lot of data to start with, and it also means that the product works best in areas where there is high population density that we have a lot of people whose data we can have. The second thing with mobile network data is that the locations are not accurate, because the data that we get is the mobile devices connected to so called “Mobile Cells” or like network cells... Well, we don’t know the exact GPS location of the person...”*

4.3. Privacy issues

Live and continuous data streams from cellular towers or base stations are critical components for the delivery of mobile services by network operators. These data streams include information about every mobile phone user connected to the network. A Senior Data

Scientist at Telenor Norway stated: *“If you use your cellphone today, then you basically just have to accept that this kind of data is collected.”* Despite this, the process of collecting and analyzing mobile data is still done according to strict privacy regulations. We asked the Head of Data Insights from Telia Sweden whether this process presents surveillance. He emphasized that the company uses aggregate rather than individual data to develop generic insights, he explained: *“...to me when you say surveillance that means you’re tracking someone. And the term surveillance from a legal standpoint is also not optimal because it can mean something else. So, when we provide these insights, we cannot see individuals, it is anonymous, it is aggregated information. So, we’re not surveilling anyone. There is no surveillance here, there are no individuals in here. What we are providing is general insights on larger movements of people in society.”*

A legal viewpoint on this was offered by a Data Protection Officer from Telia Sweden who explained that telecom operators have to commit to telecom privacy laws, she said: *“...how the data can be anonymized so that we can create meaningful statistics to show people movements but at the same time no individual is identifiable within the data. We have created a process that, well, processes the data from the network and remove the direct identifiers, and group the data to a certain group size so we basically know that there is a group of people in this location at a given time, but we don’t know who the people are. And if the groups are too small, if that’s only one person in a certain location, then that data will be deleted, and we don’t have that location at all. So, the grouping (aggregation) is the key”*. Still, the Head of Method and Development at NIPH explained that sometimes people’s movements across municipalities can be limited and even data aggregation would not help in protecting their privacy which raises concerns about how it can be used. She said: *“...you will also have issues there because let’s say we take municipality in the very south end of Norway, and you take a small municipality in the northern part of Norway; the amount of people that travel between these two municipalities is very small, so also there you can enter some kind of privacy issues.”*

In general, whether these measures by telecom operators are sufficient to protect individual privacy while leveraging the potential of mobile data is a critical issue for all actors involved especially mobile phone users. Our survey data, for instance, show that more than 60% of our Danish respondents did not trust their authorities and more than 50% of our Swedish respondents reported the same when it comes to using mobile data to understand population movements. The rest of the Nordics seem to trust their authorities in more than 70% of the respondents. Furthermore, while

telecom companies emphasize that generating patterns of population movements is done using aggregated data based on groups of individuals rather than an individual, about 50% of our respondents believe that such patterns are analysed with identifiable personal data, with 20% of them strongly believing this likelihood.

5. General discussion

The use of mobile data during disease outbreaks is a novel phenomenon and a new data source to support epidemiological work. While public health surveillance that is based on the collection, reporting, and analysis of data from multiple sources is still prevalent [8,11], there is increasing recognition of the potential of mobile data to help in understanding disease dynamics especially disease transmission and control [5,18,19]. Our investigation of the use of mobile data in a Nordic context shows that there are a number of potential application areas of this data which are mostly focused on understanding patterns of population movements. Population movement patterns describe how groups of individuals move or travel from one location to another and include details about geographical areas, distance, time, number of trips, and statistics of generic changes and developments of movements. These patterns are developed using data aggregation and analysis of raw mobile data generated from cellular communications to produce origin-destination matrices. Data aggregation is a process of compiling data to develop generic and abstract patterns of movements in an anonymized way. It is an essential censoring practice that involves anonymization measures such as time-delays, location approximation, and group rather than individual base numbers. This study shows that data aggregation is essential for using mobile data in a disease outbreak which is consistent with earlier research [16] on the usefulness of aggregate descriptions to understand human behavior and obtain usable information from massive amounts of data such as mobile data.

There are two main application areas of mobile data which are characterized by categories of use for epidemiological work in the context of the Covid-19 outbreak. First, using aggregated mobile-based data for informational and validation purposes. The interview data show that public health authorities were provided with access to interactive maps and visualizations that show patterns of population movements in terms of total number of trips on a certain day or week, distances travelled, travel times, travel routes, and general changes in travel behaviour (i.e. during Easter 2020 Sweden witnessed 40% decrease in travel compared to the previous year). Public health authorities were able to use informational data for a variety of purposes such as understanding and validating commitment to public

health measures as well as providing insights into potential risk areas where the rate of travel and movement is high. Second, using aggregated mobile-based data as data complements for epidemiological modelling and the development of prediction models. While it is not yet clear how to use mobile data by epidemiologists, our study shows insights into using this data as complements for existing data (i.e. demographic data) obtained by public health authorities from multiple sources. This involves using detailed information from mobile data about how people move to complement predictive models to make short-term predictions about disease transmission activity and develop situational awareness of the impact of public measures such as lockdowns and school closures. In Norway, NIPH used mobile data for the development of predictive models used for purposes such as the municipal population model and the individual based model. In Sweden, certain parameters (i.e. contact intensity) were used to complement the development of epidemiological models using insights from mobile data.

While using mobile data as validation and complementary data may offer novel opportunities for epidemiologists during disease outbreaks, the fact that they are subject to aggregation may result in limitations for their potential. The interview data show that limitations are related to data accuracy which can be compromised because of relying on mobile cells and time delays for anonymity purposes. For instance, the exact location of an individual is approximated making it hard to achieve accuracy in understanding spatial dynamics of disease transmission. There are also limitations in the richness of mobile data since personal information such as demographics about the individuals, which are critical in epidemiological modelling, cannot be collected or aggregated due to privacy rules. The telecom companies providing aggregated datasets from their mobile data to public health authorities do not know themselves such information, especially that this kind of personal information is strictly regulated internally as well as by GDPR. When we asked our survey respondents if they would accept relaxing EU and national privacy regulations to combat the Covid-19 disease outbreak, ca. 60% were against the idea, ca. 32% were pro the idea, whereas 8% said in certain situations, such as: “if Covid-19 becomes worse”, “in scientific(statistical) proven hotspots that can prevent the further harm and spread”, “in extreme, dire situations”, “only when people have given informed consent”, “In a situation where the health care system can't keep up” and related concerns. Further, it is important to acknowledge that mobile data is just one source of data among many other critical data sources used in epidemiological modelling. Mobile data may be useful as complements to other

types of data which limits their scope in epidemiological work. In general, the problem of generalization and abstraction in data aggregation is not particular to using mobile data in disease outbreaks, but rather it is a common problem in big data aggregation [16].

Finally, these limitations cannot be separated from concerns about individual privacy. In fact, data aggregation is essentially done to protect the anonymity and privacy of individuals besides generating meaningful outcomes of massive amounts of data. It is therefore reasonable to argue that privacy concerns cause many of the limitations on the use of mobile data in disease outbreaks despite their potential. Our survey data show that while people generally think that using mobile data to understand population movement patterns may be beneficial to public health organizations, the majority still believes that the organizations involved in data collection and analysis cannot guarantee their individual privacy. This is despite the fact that more than half of our respondents, regardless of their nationality (ca. 53%), trust the authorities when it comes to obtaining patterns of population movements using mobile data, while the rest remain skeptical about such practices.

6. Implications for practice

The analysis and discussion of our empirical data offers a number of implications for practice which can inform action to use mobile data for understanding population movement patterns during the Covid-19 outbreak. These include sharing resources and expertise between health agencies and telecom companies, extended collaboration with multiple network operators, cross-disciplinary collaboration among multiple parties, developing data and privacy guidelines, and developing novel methods and tools to address the trade-off between maintaining individual privacy and obtaining rich and detailed information from mobile data. Each of these implications is discussed below.

While our study shows that public health organizations recognize the potential of using mobile data in pandemic situations, there is lack of understanding how to leverage this data and generate value from it. This calls for quick action for sharing both resources and expertise between health organizations and telecom operators to address the urgent need of understanding disease transmission dynamics. Also, while mobile communication produce massive amounts of mobile data [1,5], usually network operators have a certain share of the mobile market which may cause limitations in the spatial and temporal outreach of the data. This may suggest that data about wider population movement patterns can be limited and there might be a need for an extended collaboration with multiple

operators at the national level in order to access more geographical areas. The ability to understand wider population movement patterns may facilitate quick action and effective mitigation and control measures. This is especially true when the number of cases is low, which is critical in controlling and stopping disease transmission. In the longer term, this might help with preparedness measures in early stages of the pandemic.

In addition, leveraging the full potential of mobile data may require both public health organizations and mobile network operators to develop guidelines regarding potential data and user privacy issues. The lack of guidelines on privacy issues can be problematic as reported earlier in literature [5,18] due to limitations in using the data because of the trade-off between maintaining people's privacy and obtaining detailed and meaningful information about population movement patterns from the data. This urges action by policy makers to enact measures for sharing effective datasets with public health organizations without compromising individual privacy. It also urges action by data experts to develop tools that can reduce the level of abstraction for better representation of reality.

7. Conclusions and future research

This study investigates the use of mobile data for understanding patterns of population movements in relation to Covid-19 transmission. The use of mobile data in a global pandemic situation has rarely been studied before due to its novelty and suddenness. The current study contributes with empirical insights into their use by public health agencies during the Covid-19 outbreak in Sweden and Norway. These insights show that mobile data is a novel data source that can be used by epidemiologists for validation and complementary purposes. This is one important contribution which addresses the first research question. Using mobile data in complementary ways can help in offering opportunities for better development of epidemiological models and tools that use timely information which is critical for better analysis and prediction of disease transmission. Also, the current study contributes with a number of key implications for practice which are aimed at both immediate and future actions. These implications are focused on developing strategies for using mobile data through cross-disciplinary collaborations among data experts, epidemiologists and privacy officers. They are also focused on developing guidelines to address privacy matters which have been found to affect the extent to which mobile data can be useful for epidemiological work. In addition, it is important to mention that these implications are observed in an environment where laws and regulations makes it hard to implement privacy-invading tracking of

individuals. We believe that they can be relevant to inform health authorities in other contexts with a similar regulatory environment. We therefore suggest that the development of regulatory and privacy frameworks, the design of tools and methods for the analysis of effective and meaningful abstractions from massive amounts of data, and the application of mobile data in developing predictive epidemiological models can all be key areas for a future research agenda.

8. References

- [1] B. Baesens, R. Bapna, J. Marsden, J. Vanthienen, and J. Zhao, "Transformational issues of big data and analytics in networked business". *MIS Quarterly*, vol.40, 2020.
- [2] Sh. Zuboff, "The Age of Surveillance Capitalism", Profile Books Ltd, London, 2019.
- [3] S. Vodanovich, D. Sundaram, and M. Myers, "Research commentary—digital natives and ubiquitous information systems". *ISR*, vol.21, no. 4, 2010.
- [4] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, and P. Niu, "A novel coronavirus from patients with pneumonia in China". *New England Journal of Medicine*, 2020.
- [5] A. Wesolowski, C.O. Buckee, K. Engø-Monsen, and C.J. Metcalf, "Connecting mobility to infectious diseases: the promise and limits of mobile phone data". *The Journal of Infectious Diseases*, vol. 214, suppl. 4, 2016, S414-S420.
- [6] C. Fraser, D.A. Cummings, D. Klinkenberg, D.S. Burke, and N.M. Ferguson, "Influenza transmission in households during the 1918 pandemic". *American J. of Epidemiology*, vol.174, no.5, 2011, pp.505-514.
- [7] C.J. Duncan, and S. Scott, "What caused the black death?" *Postgraduate Medical Journal*, vol.81, no. 955, 2005.
- [8] S. Briand, A. Mounts, and M. Chamberland, "Challenges of global surveillance during an influenza pandemic", *Public Health*, vol.125, no.5, 2011, pp. 247-256.
- [9] D.L. Heymann, and G.R. Rodier, "Global surveillance of communicable diseases". *Emerging Infectious Diseases*, vol.4, no.3, 1998, p.362.
- [10] A.D. Langmuir, "Communicable disease surveillance: evolution of the concept of surveillance in the United States". November 12, 1971, pp. 681-684
- [11] S.B. Thacker, and D.F. Stroup, "Future directions for comprehensive public health surveillance and health information systems in the United States". *American J. of Epidemiology*, vol.140, no. 5, 1994, pp. 383-397.
- [12] N. Cox, T. Brammer, and H. Regnery, "Influenza: global surveillance for epidemic and pandemic variants", *European J. of Epidemiology*, vol.10, no.4, 1994.
- [13] WHO, Interim Surveillance Recommendations for Human For Novel Coronavirus, 2013.
- [14] D. Broniatowski, M. Paul, and M. Dredze, "National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic". *PloS One*, vol.8, no.12, 2013.
- [15] A. Abbasi, S. Sarker, and R. Chiang, "Big data research in information systems: Toward an inclusive research agenda", *Journal of the AIS*, vol.17, no.2, 2016.
- [16] I. Constantiou, and J. Kallinikos, "New games, new rules: big data and the changing context of strategy", *Journal of Information Technology*, vol.30, no.1, 2015, pp. 44-57.
- [17] J. Jiang, and L. Luo, "Influence of population mobility on the novel coronavirus disease (COVID-19) epidemic: based on panel data from Hubei, China". *Global Health Research and Policy*, vol. 5, 2020, pp.1-10.
- [18] J. Cinnamon, S. Jones, and W. Adger, W. "Evidence and future potential of mobile phone data for disease disaster management", *Geoforum*, vol.75, 2016, pp. 253-264.
- [19] A. Wesolowski, T. Qureshi, M.F. Boni, P.R., Sundsøy, M.A., Johansson, S.B. Rasheed, K. Engø-Monsen, and C.O. Buckee, C.O "Impact of human mobility on the emergence of dengue epidemics in Pakistan". *Proceedings of the National Academy of Sciences*, vol.112, no. 38, 2015, pp. 11887-11892.
- [20] A. F. Cooper, J. J. Kirton, and T. Schreker (eds), "Governing Global Health". Routledge, London, 2016.
- [21] T.C. Reluga, "Game theory of social distancing in response to an epidemic", *PLoS Computational Biology*, vol.6, no. 5, 2010.
- [22] P. Ågerfalk, K. Conboy, and M. Myers, "Information systems in the age of pandemics: COVID-19 and beyond", *European J. of Information Systems*, 2020.
- [23] E. Nsoesie, M. Marathe, and J. Brownstein, "Forecasting peaks of seasonal influenza epidemics". *PLoS currents*, vol. 5, 2013.
- [24] K.T. Eames, and M.J. Keeling, "Contact tracing and disease control", *Proceedings of the Royal Society of London*, vol.270, no. 1533, 2003, pp. 2565-2571.
- [25] D. Klinkenberg, C. Fraser, and H. Heesterbeek, "The effectiveness of contact tracing in emerging epidemics", *PloS One*, vol.1, no.1, 2006.
- [26] W. Wang, L. Ying, and J. Zhang, "On the Relation Between Identifiability, Differential Privacy, and Mutual-Information Privacy", *IEEE Transactions on Information Theory*, vol.62, no. 9, 2016, 5018-5029.
- [27] A. Acquisti, L. Barndimarte, and G. Loewenstein, "Privacy and human behaviour in the age of information", *Science*, vol.347, 2015, pp. 509 - 514.
- [28] P.E., Ketelaar, and M. van Balen, "The smartphone as your follower: The role of smartphone literacy in the relation between privacy concerns, attitude and behavior towards phone-embedded tracking". *Computers in Human Behavior*, vol.78, 2018, pp. 174-182.
- [30] U. Schultze, and M. Avital, "Designing interviews to generate rich data for information systems research". *Information and organization*, vol.21, 2011, pp.1-16.
- [31] M.Q. Patton, "Qualitative research and methods: Integrating theory and practice. SAGE Publications, Thousand Oaks, CA, 2015.
- [32] R. A. Kleinman and C. Merkel. Digital contact tracing for COVID-19. *CMAJ*, vol 192, no. 24, 2020.
- [33] K. Riemer, R. Ciriello, S. Peter, and D. Schlagwein. Digital contact-tracing adoption in the COVID-19 pandemic: IT governance for collective action at the societal level. *EJIS*, vol 29, no. 6, 2020.
- [34] A. Urbaczewski, and Y.J. Lee, Information technology and the pandemic: A preliminary multinational analysis of the impact of mobile tracking technology on the COVID-19 contagion control. *EJIS*, vol. 29, no. 4, 2020.